TS-EAS EAC-CPF Berlin meeting

Attendees (in person)

Name	**Institution**	**Role**
Caitlin Rizzo	Penn State University, US	EAC-CPF team
Cory Nimer	Brigham Young University, US	Documentation team
Gerhard Müller	Staatsbibliothek zu Berlin, DE	TS-EAS ex officio
Joost van Koutrik	Het Utrechts Archief, NL	TS-EAS
Karin Bredenberg	Kommunalförbundet Sydarkivera, SE	TS-EAS co-chair
Kerstin Arnold	Archives Portal Europe Foundation, EU	EAD team lead
Regine Heberlein	Princeton University, US	former EAC-CPF team member
Silke Jagodzinski	Geheimes Staatsarchiv Preußischer Kulturbesitz, DE	EAC-CPF team lead

Monday, 9 March 2020

Welcome and introduction

Welcome by Herrn Kriese (Deputy Director, Geheimes Staatsarchiv, Preußischer Kulturbesitz)

Introduction to the meeting by Silke

Organisational matters

- Notetakers:
  - Monday - Regine
  - Tuesday - Cory, Joost
  - Wednesday - Silke/Karin, Regine, Cory
  - Thursday - Joost
- Live note taking during the meeting -> aim to have final notes by end of March 2020
- Report for website / Standards Committee - Kerstin
- Introduction of participants in the room
- Wrap-up session each day -> Zoom meeting options for all TS-EAS members to join sent by Karin in advance
- Sessions with online presenters:
  - Assertion description - Mark
  - RiC - Florence
  - Schema design - Mark
  - Documentation - Ailie

Objectives of the meeting (decide on)

- Agree and finalise some long time discussed questions:
  - Identifiers
  - Single elements and attributes
  - EAD3 reconciliation
  - Namespaces (XML / Xlink)
  - Documentation

Scope of the EAC-CPF Revision

- Bug fixing, even currently are no bugs reported
- User request, see GitHub issues
- EAD3 Reconciliation
- Two-tier strategy: released schema version 2010 revised end of 2018
  - Minor changes - existing EAC-CPF files are still valid
  - See homepage for details
  - Schema, RNG, Tag Library released
- Next phase for major revision with 6 main topics: names, dates, identifiers, relations, assertion description and EAD3 reconciliation

- Names, dates - already discussed and finalised during the Austin meeting, cf. 20190801_EAC-CPF_minutes.pdf
- Identifiers, EAD3 reconciliation, relations, assertion description - to be discussed and hopefully finalised in Berlin

Timeline

- Planned to finish discussion and decisions latest in Chicago, August 2020
  - Full day meeting in Chicago planned
- Work on Schema and Documentation after Chicago, August 2020
- Release for call for comments in autumn 2020
- Release new version early 2021
- Continue with monthly virtual meeting, 90 minutes
  - Follow in GitHub: notes and issues
  - Work with GitHub: one issue for each element/attribute to facilitate schema and documentation work and to provide a summary for all decisions

Status of EAC-CPF revision

Namespaces

- Agreed to remove Xlink namespace
- Open question to remove XML namespace
  - Discussion expected on Tuesday

Control: align with EAD3 as much as useful and possible

- Discussion expected on Tuesday and Wednesday
- Question of spelling names (camel case, lower case...) - to be decided
- Agreed to add optional attributes, aligned with EAD3:
  - @audience: to all elements
  - @endcodinganalog: all or only with crosswalks?
  - @relatedencoding: all
  - @countryencoding: // control
  - @dateencoding: // control
  - @langencoding: // control
  - @repositoryencoding: // control
  - @scriptencoding: // control
  - @countrycode: //mainetananceAgency
  - @instanceurl: //recordId
- Agreed to not to use @localType in // control
- In favour to add language attribute to all (non-empty) EAC-CPF elements
  - Also @script
  - Also @transliteration
  - Pending: decide according elements
- //eventType: element content becomes attribute value
  - Add new mandatory attribute @value with restricted values: cancelled, created, deleted, derived, revised, unknown, updated

- //agentType: element content becomes attribute value
  - Add new mandatory attribute @value with restricted values (human, machine, unknown)

CPF Description

- Agreed to add optional attributes to //list
  - @listtype
  - @mark
  - @numeration

Plural/singular elements: handle in a distinct way

There are formal and informal descriptive elements within //description. Formal descriptive elements can be bundled within their plural element and they can also be repeated as singular elements, e.g.:

```
<places>
  <place/>
  <place/>
  <place/>
</places>
```

vs.

```
<place/>
<place/>
<place/>
```

Informal descriptive elements are <structureOrGenealogy>, <generalContext>, <biogHist>. They are repeatable but can't be bundled in a plural element.

The current situation for the formal elements bears complexity, because there are multiple options to process these (repeated) elements. The proposed solution uses singular elements only in plural elements and makes only the singular elements repeatable. This makes processing easier as well as as there will always be a bundling element - if used at all - including one or more singular elements, e.g. <functions> (0..1) – <function> (1..n) – <term> (1..n) with @lang and other child elements of singular/plural elements.

Identity

There were user requests to rework on identity elements, c.f. GitHub issues #44 and #55. It was already decided to rename <entityId> to <identityId>, c.f. <entityId> #115.

It was also already decided to change the encoding of the entity type by adding a new parent element <entityTypeDeclaration>. Child elements would be a mandatory, non-repeatable <entityType> to declare one of the three entity types, person, corporate body, or family. One or more optional child elements <localEntityType> would contain any text to specify the entity type, if needed. The optional attribute @localType should be added to both elements.

A lively discussion started around the element <entityType>.

- Kerstin pointed out that similar elements, like <eventType> and <agentType>, elements with a fixed list of terms as content, were turned to use these fixed values with a new, mandatory attribute @value instead. Furthermore, a “declaration”

wrapper element is only used within //control so far and might be avoided in the description area.
- Gerhard and Regine suggested to spare the additional element <localEntityType> and to make <entityType> repeatable, though that came with the difficulty that it might compromise interoperability and that having an option for local types was the initial user request. Alternatively, Gerhard suggested to make @entityType an attribute of <identity>.
- As EAC-CPF is following ISAAR(CPF) we are going to keep the limited list of entity types (CPF), as Karin indicated.

Decision: Kerstin creates a paper on Topic: Entity types to prepare a discussion and decision on this encoding. The discussion might be started in one of the virtual meetings until August 2020.

Names

A new encoding of names was agreed in and following the Austin meeting to simplify the encoding, c.f. paper on Topic: Names.

- The element <nameEntryParallel> is renamed to <nameEntrySet>.
- The element <authorizedForm> is removed and replaced by the new attribute @status with values authorized and alternative.
- The element <preferredForm> is removed and replaced by the attribute @preferredForm with values true and false.

Do we need to refer from a name entry to a convention and to a source? If yes, the solution for assertion description, see below, should be used.

Dates

In general we agreed to keep the current date encoding as there is no user request to change, but we will look at dates in terms of the EAD3 Reconciliation.

We decided to use the EDTF rules to encode open, uncertain and unknown dates, c.f. paper on Topic: Dates.

Status of EAF

Joost reported about the status of an EAC-F/EAF schema to encode the description of functions, following ISDF.

Topic: Assertion Description

Mark provided a new draft on Including Evidence-Based Assertions in Encoded Archival Standards and presented his proposal online. After some discussion during the virtual meetings and further comments in the online document, the final proposal was accepted by the attendees now.

Decision: The solution proposed in the current paper on Topic: Assertion was accepted:

1. Add a new optional element <citedRange> as a child of the existing element <source>.
2. Add new optional attributes @sourceLink and @maintenanceEventLink.

Remaining open questions:

1. Decision on elements where the new attributes @sourceLink and @maintenanceEventLink have to be added.
2. Decision if a similar link to the element <conventionDeclaration> is needed.
3. Replace <sourceEntry> with <citation> within <source>.

Task, also related to Topic: Relation and EAD3 Reconciliation, is to review the linking attributes in EAC-CPF and consider a “naming policy”.

Single issues

The topic of Single issues to decide was postponed. Most of the issues can be discussed and decided in virtual meetings or in the context of EAD3 Reconciliation.

Wrap-up and summary

Daily summary for TS-EAS members in a virtual session.

Tuesday, 10 March 2020

Topic: Relations

Gerhards starts with an introduction, which was based on his key issue paper Topic: Relations concerning the revision of encoding relationships within EAC-CPF. His main objectives are:

- simplification and clarification of encoding
- ensuring cross-domain interoperability
- reconciliation with EAD3

Short introduction to SoNAR

A very early visualization of concepts elaborated by the project SoNAR (IDH) that has been funded by the German Research Foundation was demonstrated. The visualizations included geographical and occupational distributions of a sample of 3,068 agents linked with archival records encoded with EAD. Kalliope encodes relations between archival resources and agents that created, received, or are subject by using authority records with shared persistent identifiers. Kalliope reuses authority records by transforming data from MARC21 to EAC-CPF.

Authority records about an agent may include explicit relations between the agent and

A. persons (e.g. family),
B. corporate bodies (e.g. affiliations),
C. conferences,
D. time (e.g. birth date),
E. occupation (e.g. a profession a person is identified with), and
F. places (e.g. headquarter of a corporate body)

these related entity types are, whenever possible, identified by persistent identifiers of shared authority records. The Integrated Authority File of the German National Library has an ontology available in OWL online to support linking and sharing data.

EAD in turn, has explicit relations between a record resource and i.a. an agent encoded (<controlaccess> sub-elements with @role, @source, @authfilenumber, @normal). Information about a repository is also recorded in sub-elements of the <repository> field using @role, @encodinganalog, @source, and @authfilenumber (for ISIL-Code). @encodinganalog is generally used for an internal database number, while the @authfilenumber is used for authority file numbers, e.g.

```
<persname role="Verfasser" encodinganalog="DE-611-PS-174638”
source="GND" authfilenumber="11851508X" normal="Bretano,
Sophie">Brenato, Sophie (1776-1800)</persname>
```

Relationship types in the context of archival description

EAC-CPF and EAD3 currently support the encoding of data about relations between the following entity types:

EAC-CPF	**EAD3**
CPF to CPF	Resource to Resource
CPF to Function	Resource to Function
CPF to Resource	Resource to CPF
CPF to Time	Resource to Time
CPF to Place	Resource to Place
CPF to Subject (Mandate, Function, Occupation, Legal Status)	Resource to Subject (Subject, Function, Occupation, Legal Status, Works)
CPF to Local Description	Resource to Works (Monographs/Periodicals) (bibliography)
	Resource to Instance (dao)

- CPF to CPF and Resource to CPF in this context also include agents that are responsible for maintaining records (holding organizations)

Not available are events (e.g. marriage), where the event itself is identified with a Place, a Time, or other related entities. EAD3 has the Chronology Item Set <chronitemset>, which was introduced to allow for bundling several places and several events to the same date field. It includes the element <event>, which, however, encodes unstructured text only.

The scope of modeling events with the Encoded Archival Standards should be addressed at the Chicago meeting.

In order to encode data about relations a set of elements is needed:

- Name of the related entity (required)
- Type of related entity
- Type of relationship
- Role of entity in relationship
    - Characterizes the context of each of two entities in a relation (e.g. if type is hierarchical, one entity is the superior and the other the subordinate)
- Time or time frame of a relationship
- Place of a relationship
- Source of the relationship statement

Relationships currently can be recorded using unstructured text, structured text and URIs. It would be preferable to have a method of differentiating between these approaches so that it can be clear that URIs are available.

Relationships in EAC-CPF descriptions section are generally straightforward by sharing the same child elements and attributes. Only the Place <place> uses different child elements.

When recording relationships, we need to be aware of the depth of information available. Generally these are recorded in simple fashion A → B. Additional information about the relationship type (e.g., not just "associative", but "friend") might be difficult to obtain or uncertain/unconfirmed. In some cases, particularly for events, there may also be an associated date and place for the relationship.

We need to look at the relation section, perhaps decluttering in the descriptive section of EAC-CPF, to increase use of structured relationships section. Whereas the descriptive section could concentrate on simple relationships, ideally with URIs, but also allowing for literals, the relations section could concentrate on describing the relationship itself. Here again the values can be both available as literal only or as resource represented by URI.

It would depend on the level of detail, where a relationship is recorded: when only few data are available it might make more sense to have it in the descriptive section, while in other cases relations can be described in detail in the relations section.

Both quality of data and level of detail should be in tag library examples.

The proposal calls for replacing current relations child elements of <relations> (<cpfRelation>, <resourceRelation>, <functionRelation>) by a unifying relations child element <relation>. A first intention was to reuse following child elements: <relationEntry>, <date>, <dateRange>, <dateSet> and to introduce new child elements: <relationType>, <relationRole>, <place> and <assertion>.

Proposal

To **unify** the relations child element by using a **<relation>** element, similiar or other singular/plural elements was agreed.

By solving the assertion description with another solution, the element for assertion or evidence wouldn't be used. Instead the new attributes to link to a source, **@sourcelink**, and to a maintenance event, **@maintenanceenventlink**, will be used within <relation>. The exact attribute names might be changed.

The element **<relationEntry>** is used to identify the related entity with a text. It was agreed to **rename** the element **to <targetentity>** in order to be precise in the elements name.

An attribute for the type of the target entity must be used with limited values, i.e. **@targettype** with values corporatebody, person, family, resource and function, within <targetentity>. This attribute corresponds to the EAD3 attribute @relationtype.
This approach allows to extend the entity types, but also simplifies the encoding. New entity types might be added by including the Records in Contexts approach in the EAC-CPF schema or by introducing a new entity description for events.

In order to encode the source of the related entity and the value by URIs, 3 **new optional attributes** should be introduced to encode relations:

- **@vocabularysource**: text or literal, where the related entity comes from,
- **@vocabularysourceuri**: URI of the source, and
- **@valueuri**: URI of the related or target entity at the mentioned source.

The attribute names might be changed later on.

The element <targetentity> does not contain text or mixed content, but follows the name entry approach with a mandatory child element **<part>**. The <part> element contains text and is repeatable.

The new optional element **<relationtype>** is used to encode the type of a relation, next to the attribute @targettype within <targetentity>. The element contains text and is repeatable. 3 new optional attributes are used to encode URI of the elements content. The element <relationtype> might contain local terms, terms in different languages or other text to provide a description of the relation type.

A new optional element **<targetrole>** is used to specify the role of the related entity within the relation. It is a repeatable text element with optional attributes to encode the URIs of the elements content.

Next to these elements, the existing optional child elements for dates (**<date>**, **<dateRange>**, **<dateSet>**) will be kept.

The distinct encoding of a **place** connected to the related entity wasn't discussed during the meeting and will be treated during a virtual team meeting.

Other existing elements within the relations part (<descriptiveNote>, <objectBinWrap>, <objectXMLWrap>) in the current schema haven't been part of the new proposal. Their usage will be discussed during a virtual team meeting.

Example encoding:

```
<relations>
       <relation id="1" sourceLink="#43" maintenanceLink="#43">
         <targetEntity targetType="person" vocabularySource="GND"
                     vocabularySourceURI="http://d-nb.info/gnd/"
                                valueURI="http://d-
nb.info/gnd/1190671301">
                <part localType="family">Arendt</part>
                <part localType="given">Paul</part>
                <part localType="dates">1873-1913</part>
```

```
            </targetEntity>
            <relationType vocabularySource="GND"
                        vocabularySourceURI="https://d-
  nb.info/standards/elementset/gnd#"                  valueURI="https://d-
  nb.info/standards/elementset/gnd#Family"
         localType="text" lang="en">Family</relationType>
            <relationType vocabularySource="GND"
                        vocabularySourceURI="https://d-
  nb.info/standards/elementset/gnd#"                  valueURI="https://d-
  nb.info/standards/elementset/gnd#Family"
         localType="code">bezf</relationType>
            <targetRole vocabularySource="schema.org"
                        vocabularySourceURI="http://schema.org"
                              valueURI="http://schema.org/parent"
  localType="text"
         lang="en">parent</targetRole">
            <targetRole vocabularySource="schema.org"
                        vocabularySourceURI="http://schema.org"
                              valueURI="http://schema.org/parent"

         localType="code">par</targetRole>
            <date>...</date>
            <place>...</place>
            <descriptiveNote>...</descriptiveNote>
         </relation>
</relations>
```

[Additional examples of resource, etc. needed to complete proposal.]

Decision:

1. Proposal as described above will be distributed with the call for comments on the draft revision, rather than being sent out to the larger committee beforehand.
2. Rename <relationEntry> to <targetEntity>.
3. Add element <relationType> to encode/describe the type of the relation.
4. Add element <targetRole> to encode the role of the related entity.
5. Keep date and place information, also descriptive note element.

Topic: Referencing

The topic of referencing attributes was postponed and nearly solved by decisions to use @...link attributes for assertion description and relations. A paper on this topic summarises the status.

Remaining open questions:

1. How to name the attributes: @...link or @...ref?
2. Where to add and which attributes exactly?
    a. Would we envisage having a general @target attribute as in EAD3? Or mainly context-specific targeting attributes?

RiC-CM and RiC-O

Online presentation on RiC-CM and RiC-O provided by Florence.

The slides are from the conference held at BNF in January, translated to English. The slides are available under CC BY 4.0 and can be reused with this license.

Introduction

ICA had published the 'old' standards from 1994 to 2008. There were some inconsistencies and overlaps, like the "Administrative/biographical history" element in ISAD(G) and the "History" element in ISAAR(CPF). Other problems: relations as defined in ISAAR(CPF) are not yet present in ISAD(G).

ICA decided to develop a new global standard, Records in Contexts (RiC), which had to be backwards compatible with the existing four standards. The differences are:

RiC focusses on entities and on defining them, where ISAD(G) etc. focus on description rules. The glossary in RiC is significantly smaller. In the current ICA standards glossaries (thus definitions of concepts) are present but as, partly quite short, introductory sections. RiC-CM is mainly about defining each of the entities that exist in the world of archives.

These entities are (re)defined, added properties (attributes) and relations between them. Relations are core to the model of RiC, as opposed to the description being core to ISAD(G) etc.

The four previous standards were too few-dimensional to accommodate the multiple layers of contexts (see slide 11). Defining more distinct entities in RiC as well as their relations, would make them 'easier', but also potentially more nuanced and flexible.

See the example on slide 13. This graph shows the relations focused on the president of Radio France. In the finding aid some series were described about Radio France, the mandates, functions and others, grouped together as entities and their relations. Supposedly this form might be more usable to the general public than the 'traditional' finding aids. Note that these entities have relations, *pointing to* other entities. Any of these entities can provide an access point to enter the information system.

See slide 16 about the development of RiC by EGAD. The members 'all over the world' means the Americas, Europe and Africa. One of the main differences from the previous EGAD group is that this group has to continuously maintain and develop the standard going forward.

RiC is currently comprised of three parts: Conceptual Model (CM), Ontology (O, technical specification of the CM), Application guidelines. RiC allows to see the archival world as an

oriented graph, because it consists of entities and relations. EGAD hopes that RiC-O will not be the only technical representation of RiC-CM. There will undoubtedly be other representations. According to Florence, EAD and EAC-CPF should be made compliant with RiC-CM, as RiC-CM replaces ISAD(G) and ISAAR(CPF), on which EAD and EAC-CPF are based.

EAD and EAC-CPF could therefore be viewed as a subset of the Ontology. Before RiC-O there was no internationally supported domain ontology for archives. (There were a number of local initiatives.)

Work on the Application Guidelines is still to be started. This could encompass training courses, examples, etc.

See slide 19 for the development history and roadmap. On the first version of RiC-CM, EGAD received a lot of comments. These were processed in version 0.2 of RiC-CM. Version 0.2 is to be regarded as the only leading version as of its publication.

RiC-O can now also be found on Github. Florence will keep TS-EAS updated on news in this regard. People can create questions and issues on the GitHub site.

EGAD is also preparing the full public release of RiC-CM and RiC-O version 1.0 in November at the ICA conference in Abu Dhabi. Current 0.2 CM and 0.1 O are presumably very similar to what versions 1.0 will be.

Ric-CM 0.2 has 22 entities, grouped in four hierarchies. Main difference with ISAAR(CPF): Apart from CPF, also Positions (E12): functional role of a person in a group, e.g. the president of a company; and Mechanisms (E13), e.g. systems or algorithms.

Main difference with ISDF: Activity (E15) as a broad definition, and a subcategory of an Event (E14), covering all processes, activities, functions, etc.

For Record Resources (E2), there is a distinction between individual Records (E4) and aggregations defined as Record Sets (E3). These entities have a total of 41 attributes. Some of these are repeatable or extensible. See the RiC-CM 0.2 public release document for a specification of each attribute.

There are 78 relations (and the inverse ones) in a poly-hierarchical system, see slide 29. These also have attributes, as detailed in the RiC-CM document. See representation in the chart on slides 32 and further. Here, the relations are represented two-directionally. In the next-to-last and last graphs, the relations are represented as one-dimensional.

From RiC-CM to RiC-O

See also slide 44 and on. RiC-O is a domain ontology for archives, compassing what archivists would use for describing and maintaining archives. It should be easy to maintain and be mappable with other standards. It should be immediately usable and convertible from existing metadata sets. Existing finding aids in EAD and EAC-CPF have been converted to RiC-O. Example given on slide 48 is a query of corporate bodies existing in the nineteenth century in France.

From theory to practise

The URI for the Ontology on the ICA EGAD webpage leads to a version that is readable for human readers (https://www.ica.org/standards/RiC/RiC-O_v0-1.html). The OWL-ontology is also available. There are more classes in RiC-O than there are entities and attributes in RiC-CM (see slides 55 and 57). Example given in the presentation is the relation by class, between the president of Radio France (person and position) with Radio France as a Corporate Body.

Several events have been planned in the next few months in France.

In closing, Florence suggests that EAC-CPF can be updated to be more compliant with RiC. Implementing RiC is reliant on that. In this respect, it is not about re-inventing the wheel, but keeping it moving and improving on it.

She has submitted a document with a number of ideas connected to the work being done on EAC-CPF.

Ontology and schema: usage, differences and approach: Karin Bredenberg

PREMIS Ontology and Shema

This is the international de facto standard for digital preservation, maintained by a group of experts from Europe and the US. Karin is the PREMIS chair.

The Object (see presentation) is of 4 different “types”: intellectual entity, file, bitstream and representation. The intellectual entity can be described in PREMIS as of version 3. Relations can be defined two-way, with the exclusion of Agent to Object. One of the presumptions is that data that is being maintained in other systems, should not be maintained in the PREMIS system to make it least cumbersome as possible.

PREMIS does not prescribe its implementation, but when it comes to transfer there is an XML Schema and an RDF in place to adhere to. Usage of RDF presupposes usage of a triple store. Other expressions are also acceptable, depending on which works best for the institution that’s using it. Central questions here: how is your system built, and where are you sending the information? (For different incarnations of the given example, see presentation.)

Next actions for the PREMIS agenda: Investigate mapping the RiC-O record with the PREMIS object. Seen from a RiC-perspective, there is a good match; but the PREMIS ontology was not used for RiC-O.

Question: does Karin have any experience with usage of PREMIS ontology by others?

Someone from Princeton was involved with the ontology. In most cases, the XML-expression

For BNF: they have a huge triple store, but other institutions typically do not.

Discussion (impression)

There is some hesitation behind remodelling the EAS to the current version of RiC, mostly because the current model is still a ‘moving target’ and some of the elements (notably Events/Activities) are considerably different from the previous standards and previous

version of RiC-CM. It is however the view of EGAD that the existing schemas based on ISAD and ISAAR should in future be compliant with RiC-CM.

RiC-CM will be the new guideline from EGAD for the archival community. EGAD provides the community with RiC-O an OWL Ontology for implementing it. On the other hand, there are the current EAS-standards.

It could be crucial to align the current EAD and EAC-CPF to RiC-CM. This is not a question of whether we should, but more one of timing. What does 'align' mean? One of the strengths of RiC is that it is object oriented. The remodelling of EAS to be less description oriented to more object and relation oriented in itself, makes it more aligned to the modular approach as defined in RiC-CM. These standards need not compete. Where is RiC-O positioned, as an instantiation of RiC-CM; and the EAS will become another? Or as the only instantiation?

The current approach as discussed this morning moves to the object approach as espoused by RiC-CM. Current problem is that it is still under development, and the current planning for the revision of EAC-CPF is by late 2020. The approach as of now is to make EAS more modular.

One of the other issues is that RiC-CM and RiC-O are developed and will presumably be maintained with ICA funding. The EAS is mostly self-funded by its members. It is even difficult for the EAS team to get answers from ICA to its questions. This is an issue that needs to be resolved between SAA and ICA.

TS-EAS: schema development and shared schema

Mark explained the current development of the shared schema. The current idea is that we will move forward with a single GitHub repository that will be able to generate EAC-CPF and EAD3 as well as any additional EAS in future. With this approach, we will still be able to retain differences between the schema deliverables (e.g. filedesc in EAD3 but not EAC), if desired and when required, but we will be concentrate on specifying what the differences are up front, rather than being forced to constantly review and attempt to align any drift in definitions between the different driver schemas, which is far more complex as the current process shows.

A very important decision that we still need to make: do we want to retain separate namespaces for EAD3 2.0 and EAC-CPF 2.0, or do we want a single namespace for the EAS standards. There are advantages and disadvantages to both options. The biggest advantage is that we wouldn't need to create new namespaces for any additional EAS schemas (but if there are no others, then there is no advantage). The biggest disadvantage is that we would need to come up with yet another namespace if we go with the single EAS namespace.

A number of issues are relevant here:

Schema principles are still in draft and being worked on. These have been broken down in three main aspects:

- Schema design / technical principles;
- Tag library principles for examples and definitions;
- General standard definition principles, e.g. same name means the same model and does the same thing, don't use the same name for elements and attributes.

What is the schedule for these principles? A number of principles on schema design need to be written down and agreed upon. We do have various points in this meeting where we talk about conventions on a higher level. We should be able to have a system of quality control on decisions on EAS. A set of principles should make this control possible and easier to communicate more ‘uniformly’.

Spelling question

In EAC-CPF, CamelCasing is used to make it more userFriendly. Suggested is use of lower case going forward. There should be no ambiguity in terms when only lower case is used; so avoid creating new terms with other meaning when combining two separate terms in one element or attribute name.

Decision: The team agrees on the use of lower case.

XML namespace

First approach was to remove name spaces, as has been done in EAD3. What was the argument for removing it?
This was coming from the XLink perspective. The reason for removal in EAD3 was to not be dependent on anything. The elements that we’re using in EAC-CPF are conventions anyway.

Discussion focussed on @xml:id and @xml:lang. @xml:lang is still being used in EAC-CPF, but not anymore in EAD3. It is an obvious element in any type of XML documents; so it would be easily mapped to for example MODS or METS. It would be easier to remove XLink than XML.

We don’t know of any other standards that have removed XML name spaces. If we introduce XML to EAD3, all elements would need to be ‘changed’. And we would need to explain why the XML name spaces are being adhered to and not any others. Mark proposes to add XML base and XML name spaces. Kerstin stresses the need to be consequent about the choices being made and consequently carrying them out.

Decision: Hand a decision over to the Schema team to analyse the benefits and consequences of usage or removing of xml ns.

Wrap-up and summary

Daily summary for TS-EAS members in a virtual session.

Wednesday, 11 March 2020

EAD3 reconciliation

Kerstin is going to introduce today's topic: EAD3 Reconciliation. Caitlin and Kerstin prepared a paper on shared and similar elements and attributes in EAD3 and EAC-CPF. The aim is to ensure that shared elements, i.e. elements that use the same element names, really are of the same scope and are hence using shared definitions in both standards. Similar elements on the other hand must not use the same name and need to be defined differently.

Namespace

We already agreed to remove xlink namespace from EAC-CPF.

Keeping or removing XML namespace is not decided yet. It was introduced with EAC-CPF, i.e. EAC beta did not have it, nor did any EAD version to date. Schema team is asked to dive into this topic:

- look for examples of other schemas that removed the XML namespace or are using their own variation of XML namespace attributes (as e.g. METS does),
- look at the attributes definition to decide, if keeping the XML namespace spares us to define the attributes in questions (especially @xml:lang) within the EAS or related Schematron.

A decision should be made until our next in-person meeting in Chicago.

Language related attributes

The set of base attributes in EAD3 includes the attribute @lang that is in fact available in all elements, except <ptr>. To keep in mind for the usage of EAC-CPF decision on @(xml:)lang attribute.

The EAC-CPF Tag Library defines that @xml:lang contains values in a list taken from the IANA Registry. This list was never defined in the schema.

Decision: Language codes should be added to the schematron.

The attribute @transliteration is under discussion to keep it at all or to remove. As we have use cases in the EAC-CPF team, we agreed to keep the attribute. We could even find use cases to use different transliterations in one file (e.g. having aggregate instances or having descriptions with transliterations from different languages). Rules for transliteration should be declared via conventionDeclaration.

Decision: Keep @translation attribute as optional attribute along @(xml:)lang and @script in all non-empty elements in EAC-CPF.

Pending question: Need to decide what “non-empty” means and if there could be cases of “empty” elements, where it still makes sense to allow for language attribution, e.g. plural elements. There would ideally be the same approach taken for EAC-CPF and EAD.

Related standards via @encodinganalog

EAD3 used the attributes @relatedencoding and @encodinganalog for cross-references with related standards. @relatedencoding is available for <ead>, <control> and <archdesc>

to declare the standard referred to. @encodinganalog is then used on the single elements that have an equivalent in that standard.

At the moment, however, @encodinganalog is only available with elements that are mentioned in the crosswalks with ISAD(G) and MARC. This could be more open by adding @encodinganalog to all elements, leaving users the choice to use any standard to refer to, not only ISAAR(CPF) or ISAD(G) or MARC.

Decision: Add @encodinganalog as optional attribute with NMTOKEN in all elements in EAC-CPF.

Formatting with @altrender and @label

We already agreed to not to add both formatting attributes to EAC-CPF.

- A use case for @altrender in EAD3 is indexing titles within EAD3 at Princeton University. The noun marker in a title beginning would be marked with an @altrender.
- Kalliope uses @label for facetting information within EAD.

Decision: The EAD team will review this issue in one of their next meetings.

Dates

Some discussion about the encoding of date information, whether it was possible to implement EDTF (listed as draft on the Library of Congress website) and its status relative to the ISO 8601:2019 revisions. Additional information is needed before knowing how to proceed.

Abbreviated names

According to policies and usages for programmers it is recommended to use descriptive and non-abbreviated names.

As we already agreed to remove the camelCase in spelling the names within the schema, we agree to change the abbreviated names in EAC-CPF into descriptive names also for the already existing elements in that context.

Decision: Change abbreviated names in EAC-CPF into descriptive names.

Mixed content in the standards

EAD3 allows mixed content to various extent and on multiple elements. In the context of the EAD3 reconciliation, this specifically refers to the shared elements <abstract>, <addressline>, <citation>, <date>, <fromdate>, <todate>, <event>, <item>, <p>, and <part>.

Discussion is going around the need and usage of mixed content in the standards. Are the standards meant to encode narrative text created by ambitious archivists.

<control> and its sub-elements as case study for shared schema option

Originally developed in EAC-CPF and incorporated in EAD during the development of EAD3. Includes circa ⅔ of shared elements, making it a good element for reconciliation. Furthermore it covers quite a few general questions when looking at its sub-elements in more detail.

<control> element in itself

Main difference is the continued existence of <filedesc> in EAD3, and currently EAD3 also includes <representation>. There already exists the suggestion that add <representation> to EAC-CPF (see #9), but what to do with <filedesc>?

Some sense that there really isn't a great need for <filedesc> in EAD3 either - title is not significantly different from finding aid title, author is most likely the processor, the publisher is probably the publisher of the finding aid. There does not appear to be a significant need for <filedesc> in EAC-CPF. TS-EAS could consider moving it out of <control> in EAD3. Alternatively, could have it included as optional in the shared schema and deal with difference when publishing the respective schema files for EAC-CPF and EAD3.

<representation> in EAD3 provides a link to a transformed/deliverable version of the EAS instance. It is an optional, repeatable element with a range of attributes, including: @(xml:)id, @(xml:)lang, @script, @transliteration, @encodinganalog, and @localtype (this is one of several examples of @localtype in <control> that needs to be remediated). Pending question with regard to the linking attributes (previously XLink): which ones should be kept?

Similar to <filedesc>, <representation> does not seem to have a strong use case for EAC-CPF, but could be included as an optional element in the shared schema. As an example for a use case in EAD: in the context of Archives Portal Europe many institutions host their own finding aids, but provide an EAD file to the service for aggregation. The link to the original institution's finding aid would then provided in the <representation> element. A similar example for EAC-CPF could be coming from SNAC (e.g. https://snaccooperative.org/view/44995506, where the EAC-CPF document currently does not point back directly to this web-based representation).

Decision: Add the element <representation> as it is used and defined in EAD3 for EAC-CPF within <control>.

On the question of @localtype, the expected use with <representation> would be for specifying the format or type of representation. However, this could potentially be specified in @linkrole instead. Could be part of the more general conversation about @localtype vs context-specific @...type attributes vs simply @type.

If we are moving toward using more specific attribute values, then should we provide value lists for the use of these fields? If so, how do we deal with outliers? Also, do we actually have a list of controlled values that we want to enforce? For example, to replace @localtype in <agencycode>, then we would need an @agencycodetype (for a controlled or uncontrolled list) and an @otheragencycodetype (to cover outliers if @agencycodetype is controlled). Which approach to take (particularly whether @other...type is needed) would vary based on the element involved. Looking at a potential <agencycode> @agencycodetype again as an example, it was suggested that even having a short list of values (isil or other) could still be valuable in analyzing data.

Suggested that we move to more specific attribute values in place of @localtype, but not have them controlled. This can then be addressed in the tag library with recommendations for values, but not enforcing them via the schema. However, there was no agreement that @localtype needs to be removed.

It was pointed out that @localtype, while @localtype can be used with any content, the expectation for the use of @localtype has been that locally used value lists would be defined in the <localtypedeclaration> section within <control>.

Need to refer this back to the Schema Team to determine whether it is a design principle to not include @localtype in <control> elements. Also need some more examples to work with for different elements, which might need more specific attributes. Also need to send along to the Schema Team the option of reinstating @type element that is specific to each element, or whether the use of element-specific type attributes is a design decision for the standard. Should be possible to resolve at the May meeting of the Schema Team. Depending on this decision, the question of specific named attributes should be tackled afterwards.

Decision: Hand a decision over to the Schema team, if @localType can be used in <control> elements.

Decision: Hand a discussion on reintroducing @type over to the TS-EAS.

Beyond @localtype, the EAD3 model currently includes a range of former XLink attributes for many elements. Proposed to retain @linkrole, @linktitle, and @href, and drop the others. Will wait to see if there are concerns from the community during the standard review.

Decision: Use @linkrole, @linktitle, and @href, from former XLink attributes and drop @actuate, @arcrole, @show.

<recordId> and <otherrecordId>

Suggestion that we widen the <otherrecordid> to also capture identifiers of instances aggregated with the current one (#54).

Some questions arose surrounding the original decision to record agency IDs as a separate element, if they could be handled through an attribute instead. One possible reason for keeping it an element would be if you needed to provide attributes about/defining them, as in our @agencycodetype example.

Suggested that @valueURI, @vocabularysource, and @vocabularysourceURI be added, as optional attributes, to encode the provenance of other identifiers. Could also use MODS's @authority and @authorityURI - but preferable to reuse existing terms in our own standards, seeing that @vocabularysource attribute already exists in the EAC-CPF format.

For example:

```
<otherrecordid vocabularysource="[code value for repository]" vocabularysourceURI="[original instance repository URI]" valueURI="[URI of original instance]">
```

Currently <recordid> is text in EAD3, but specifically NMTOKEN in EAC-CPF. Felt that there was no reason for these to be different, with a preference for moving this toward NMTOKEN. Will need to consult with the EAD Team about addressing this difference.

Decision: Discuss difference btw content resp. data type in <recordId> with EAD team.

<maintenancestatus>

Reviewed the availability of the element, asking whether it should be allowed to be repeatable with language, or whether the @(xml:)lang should be removed and the element should only be used with the list of @value attributes.

Decision: Use <maintenancestatus> as an empty element without text or other content with a required attribute @value with limited values and without language attributes. Seeing that the element is part of control and not part of the descriptive section of EAC-CPF, the

necessity to have language attribution and an option for internationalisation does not seem to be that prominent. The element is mandatory and not repeatable.

<publicationstatus>

Determined to follow the same course as above, and remove the @(xml:)lang attribute and require the use of the controlled list of @value terms.

Decision: Use <publicationstatus> as an empty element without text or other content with a required attribute @value with limited values and without language attributes. Seeing that the element is part of control and not part of the descriptive section of EAC-CPF, the necessity to have language attribution and an option for internationalisation does not seem to be that prominent. The element is optional and not repeatable.

<maintenanceagency>

Element is meant to note the institution or service responsible for EAS instance. It was suggested that we widen the scope and re-use either the element itself or its sub-elements <agencyname> and <agencycode> within <conventiondeclaration> (#67) in order to state who is responsible for the maintenance of the convention. Preference would be with using the sub-elements rather than the entire <maintenanceagency> wrapper).

Ultimately, it was decided that due to lack of follow-up from the original ticket creator, it would be preferable to close the ticket without action rather than making this change.

Decision: No additional agency information in convention declaration as there is no use case available.

Wrap-up and summary

Daily summary for TS-EAS members in a virtual session.

Following wrap-up, questions came up about whether it would be possible for online participants to add questions or comments to the meeting minutes.

It was decided that the preference would be to have others comment on the issues in GitHub if they have other opinions, rather than adding comments to the minutes. The minutes should be seen just as a record of the discussion at the in-person meeting.

Thursday, 12 March 2020

Regine and Caitlin were forced to reschedule their flight to the US. Present are Silke, Kerstin, Karin, Gerhard, Joost. Cory is phoning in via Zoom.

Documentation: tag library and best practise guidelines

Ailie Smith is joining the meeting via Zoom for the Tag Library and Best Practice Guide proposals. See the Proposal document, which compares the EAD and EAC Tag Libraries. This document has seven proposals c.q. topics for discussion; these notes follow the structure of that document. The Tag Library will remain essentially the same.

Availability information in the tag library

Proposal is to detail the differences for making clear whether an element is mandatory or not. Karin agrees with the proposal, noting that this was previously a discussion for the EAD Tag Library. Silke proposes to also note the technical cardinality, e.g.: (0..1) or (0..n).

Gerhard is asking to clarify if it has been tried to express these relations as an XPath expression. Karin has tried this, but it doesn't allow for cardinality and is not accepted further on the road as a method for clarification.

The proposal is accepted as is. Karin will talk to Alex if it's possible to make the technical cardinality expressible.

Decision: Add technical cardinality next to the description in the TL.

The 'May contain' statement is not consistently stated in the EAC Tag Library. It should be more clear if a part element 'must' occur within a parent element. Best approach would be to have a separate part for expressing the availability, to make this explicit, similar to how this is done in the EAD Tag Library.

Separate parts for description and usage have been attempted. This did not work well then. If we wanted to keep them under one heading, we could consistently write them in a logical order.

Ailie works out a few examples to make the approach visible for review.

Consistency of terminology

Currently there is a difference between the Tag Libraries: mandatory vs. required and Non-repeatable vs. not repeatable. The team does not have any strong preferences. 'Required' is also used in the Schema.

It is agreed upon to change these terms to those used in the EAD Tag Library descriptions.

Decision: Adapt EAD3 solution for availability information.

Consistency of descriptions

This concerns having the same name and description between EAD and EAC Tag Libraries.

It is suggested to have a native English speaker to go through the summaries and descriptions where elements are similar between the two Tag Libraries.

Karin will provide a separate proposal on how to handle the Tag Library in GitHub.

As for the question how we would do this if we had further EAC entities added in the future, the approach would be to keep the shared elements consistently maintained, and adding specific ones for each new EAC entity.

Expression of <control> element

Currently <control> is described in detail in the introduction to the TL as well as in the elements' part of the TL. The proposal is to concentrate on the latter and to make it clearer, which elements are mandatory, e.g. by having the according statement as the very first sentence of the "Description and usage" part, and to allow for examples.

The introductory texts are useful, but should in this case be cut down to the minimal requirements for a well-formed EAC-record. There is a simplified part that can be included, leaving the detailed information on elements in the Tag Library descriptions of those elements.

Other sections in current Tag Library

It is proposed to move the detailed parts of the introduction (background, concept, structure and semantics) to the Guidelines document.

The Tag Library conventions and release notes should be in the Tag Library.

The Background and Dedication part should be placed in another part of the website and Kathy is going to be asked to update the text.

The crosswalks should be kept in the TL. EAD has a reference label to the crosswalks.

Decision: The Tag Library should contain: conventions, release notes, crosswalk. Text on background moves to the website. The glossary should be part of the Best Practise Guide.

Best Practice Guide

The document about Names also discusses Best Practices. Examples given by EAC-CPF users are useful to share. The Tag Library has some simple examples. The Best Practice Guide could contain examples written in more detail. Ailie asks if other things should be added to this guide.

It is agreed that this should be a separate document, comparable to the approach that PREMIS is taking.

Strategy for capturing Tag Library content

Have a strategy to separate descriptions and examples between the Tag Library and the Best Practice guide. Ailie will look through the issues and questions, if there are examples given in those.

It is agreed upon to use the GitHub issues to process new examples and descriptions. Adding labels and other such issues can also be done via GitHub. Ailie agrees to reset/manage new labels in GitHub.

Ask for people participating in the Call for Comments to provide TS-EAS with all the pertinent examples from practice that they have.

Kerstin notes that there are also instances of NMTOKEN being described as an element or attribute under ‘May contain’, but it is obviously not an element or attribute. Need to ensure this term is written in square brackets for the transformation not to create links. Alex has also been looking at the links in element names that have been created, but do not lead to the (part of) the name being expressed. Be careful about what you put in the text and whether this leads to unwanted links.

Furthermore, there have been two additions:

- Regarding the attributes, it would be useful to also have examples from practice for those as well.
- The values for ISO 3166 country codes are copied from the EAD page. Suggested is to link to the codes (same as languages) at the LoC website instead of maintaining it in the Tag Library. Also align the Description definition from EAD3.

There will be no need for a separate meeting with the Schema Team.
Ailie and Lina are working on the Tag Library. Silke notes that this is expected to be a quite laborious process. For EAC-CPF there were a number of people. For EAD, there were five people doing the writing work. Managing this is made easier in GitHub branch. Ailie is in the lead for this work, and should be able to call upon assistance if the work becomes more time-consuming.

Silke will call to the TS-EAS to get volunteers for assistance when needed. Plan to split out the work in specific components: descriptions, check availability information, check links, etc. Check the Documentation and Outreach team for editorial work and other ways where they can help.

We need to create a workflow for the work on TL and BP. The review should include checks on content, consistent language, working links and terminology and will be supported by the Documentation team.

Also publish the white paper on de EAC-CPF website. Another idea could be to update the bibliography to publish on the website.

Wrap-up and summary

Non-virtual wrap-up about the whole meeting.

Monday:

- Name spaces: has been moved to TS-EAS Schema team
- Spelling names: has been decided
- Entity type: discuss based on the paper prepared by Kerstin; put draft out for comments to the EAC-team and Regine and Kathy, and add comments to the GitHub issues as discussed in the EAC-team, attendees of the meeting and Kathy.
- Comments on Link or Reference attributes, and Convention Declaration (the link issue) were discussed. These can be easily discussed in the virtual meeting. Silke will ask Mark to finalise his paper, to be put forth in the April meeting.
- No solution was found for the ‘localtype or type’ question. That was put forward to the Schema team.
- Noting the authority institution for a record identifier: tendency to use the ‘Vocabulary source’ attribute. Not resolved yet is the naming, i.e. to either stick to ‘vocabulary’ or use ‘authority’ instead.

- The ObjectBinWrap or ObjectXMLWrap question in sources (should we keep this at all) would be an issue that can be resolved in the virtual meeting, following Mark's paper. Otherwise this might be a topic for the TS-EAS: the function of EAS instances should be to point to other sources; not to incorporate them. Silke will suggest to put this on the agenda for Chicago.
- Usage of 'descriptivenote' and sub-elements of plural/singular elements warrants a short overview paper on where they are used and a recommendation on what to do with them, with example samples. (Check for next meeting: all papers should have a recommendation in them, to guide the conversation.) Try to solve this in a virtual meeting in late May.

Tuesday

- The Relations Proposal was agreed upon, with a few comments. The xml-wrapper would still be available, because it is not discussed in the paper: it is a part of the above solution on the agenda for Chicago.
- The main thing to take away from the Referencing attributes was the use of targets, also discussed in the Assertion Description paper. Request to Caitlin to ensure the points mentioned in her paper have been incorporated in this version.
- The discussions on RiC-Ontology and EAS-Schema are in the notes.

Wednesday

- The EAD/EAC-reconciliation presentation covered quite some ground (see the presentation in the Wednesday folder), but some additional issues remained and were revisited:
    - agencycode (slide 87): confirmation that @status with <agencycode> would be used with the values "authorized" and "alternative"
    - p (slide110): some of these sub-elements from EAD would not be of use for EAC, and should be avoided there; needs to be reviewed and technical options need to be explored; the same applies to abstract, item and event, which are shared elements using the same content model
    - language (slide 121): align these with EAD3 (names and definitions)
    - script (slide 126): this element might be a problem for processing; furthermore, we have an element and attribute with the same name, what we would want to avoid; EAD team to discuss alternative name (e.g. writingsystem) and to put suggestion to EAC team for approval/adoption.
    - abbreviation (slides 137 to 148): abbreviation and citation are together in control. Ask Caitlin if she wants to do a paper on this for Chicago.
    - date vs datesingle (slides 173 to 181): date in EAC-CPF has never before been a known issue, but might have to be aligned with EAD3. Wait for the discussion on this in June
    - daterange (slide 186): should this have a @status attribute ("unknownstart", "unknownend")? This will be taken to the April meeting

- eventtype (slide 217): should follow the same content model as maintenancestatus and publicationstatus, i.e. should be an empty element with no text
- eventdatetime (slide 222): EAD3 currently defines the date variants in the schema directly; decision pending to either remove it from EAD3 or add it for EAC-CPF
- agenttype (slide 229): should follow the same content model as maintenancestatus and publicationstatus, i.e. should be an empty element with no text
- sources with sub-element source (slide 243): will need to follow finalisation of the assertion description topic and based on the changes to <source> proposed in this context
- Other shared elements, not part of <control> (slide 245): address with addressline; bioghist with abstract, chronitem, chronlist, event, item, and list; function and occupation with term; legalstatus; part; possibly two discussions with bioghist (and related elements) being a possible Chicago-topic (and possibly this whole slide)
- Similar elements (slide 247)

Next steps and deadlines

Silke has created a project in GitHub for the overall release of the review and wants to add milestones for Schema, Documentation, Tag Library, Review, etc. Seems like a good way of setting up and handling a workflow; keeping in mind that most of the work will not happen in sequence. Karin will talk to Carl Wilson, OPF, who is aiding with GitHub procedures next week to see if he has any suggestions.

The old issues and comments can be closed when they’re transferred to a new one.

The next virtual meeting is on April 3rd. Silke will propose to meet jointly with the EAD team and the schema team in July 2020 to go over the reconciliation work.

Decisions

Topic	Decision	Issue
Identity: Entity type	Discussion on encoding of entity type. Kerstin creates a paper on Topic: Entity types to prepare a discussion and decision on this encoding. The discussion might be started in one of the virtual meetings until August 2020.	#131
Assertion Description	Add a new optional element <citedRange> as a child of the existing element <source>. Add new optional attributes @sourceLink and @maintenanceEventLink	#43
Spelling	The team agrees on the use of lower case for element and attribute names. Change abbreviated names in EAC-CPF into descriptive names.	
Language encoding	Language codes should be added to the schematron. Keep @translation attribute as optional attribute along @(xml:)lang and @script in all non-empty elements in EAC-CPF.	#28
EAD3 Reconciliation	Add @encodinganalog as optional attribute with NMTOKEN in all elements in EAC-CPF.	#123
	Add the element <representation> as it is used and defined in EAD3 for EAC-CPF within <control>.	#81
	Hand a decision over to the Schema team, if @localType can be used in <control> elements. Hand a discussion on reintroducing @type over to the TS-EAS.	
	Use @linkrole, @linktitle, and @href, from former XLink attributes and drop @actuate, @arcrole, @show.	
	Use <maintenancestatus> as an empty element without text or other content with a required attribute @value with	#84

	limited values and without language attributes. Seeing that the element is part of control and not part of the descriptive section of EAC-CPF, the necessity to have language attribution and an option for internationalisation does not seem to be that prominent. The element is mandatory and not repeatable.	
	Use <publicationstatus> as an empty element without text or other content with a required attribute @value with limited values and without language attributes. Seeing that the element is part of control and not part of the descriptive section of EAC-CPF, the necessity to have language attribution and an option for internationalisation does not seem to be that prominent. The element is optional and not repeatable.	#85
Topic: Names	No additional agency information in convention declaration as there is no use case available.	#67
Documentation	Add technical cardinality next to the description in the TL. Adapt EAD3 solution for availability information. The Tag Library should contain: conventions, release notes, crosswalk. Text on background moves to the website. The glossary should be part of the Best Practise Guide.	#20